Introduction to Data Analysis with Python
Cecilia Greiff
Autumn Semester 2025-2026

Trends and Cross-Country Differences in Educational Performance:
A Data Analysis of PISA Low Achievement Rates (2006-2022)

Group Project

Group 4:
Christina Falkner
Karishma Kinjal Shah
Deborah von Frankenberg

20/12/2025

1. Introduction

Described as education's "World Cup" (Coughlan, 2013), the Programme for International Student Assessment (PISA) by the Organisation for Economic Co-operation and Development (OECD) has emerged as a leading international benchmarking for evaluating and comparing national education systems. The PISA assessment has been held every three years since 2000. It evaluates students' performance across OECD countries for the three core school subjects of reading, mathematics and science. As such, it targets 15-year-old students, usually at the end of their mandatory secondary education and it aims to overall assess a country's education system (Eurostat, 2023). Countries whose students perform well on the assessment are considered to have model education systems and are consequently held in high regard by governments and policymakers worldwide. Education and participation in school life of children is essential to allow parents, including women, to work and for the development of the child itself. The OECD (2023) found that completing upper secondary education is seen as a basis for entering the labour force in many countries in addition to allowing the individual to receive a better salary and job prospects. To complete this school level, young adults need to have reached the basic skills level in reading, mathematics and science to get into upper secondary education. It was further found that progress of students in upper secondary education attainment varies greatly between countries and developed differently over time, with some countries improving greatly (OECD, 2023).

However, as per OECD (2023) data, approximately 25% of 15-year-olds in OECD member countries (including students not covered by PISA) are estimated to perform poorly across the subjects, implying the students have not yet reached Level 2 proficiency and struggle with interpreting simple texts or applying basic algorithms. This leads us to examine how low achievement rates evolved over time across countries (Research Question 1) and whether there are major differences in education performance and its development among countries (Research Question 2). It is also known that youth and schooling has been highly affected by COVID-19 as distance learning and absences of students and teachers impacted the learning of students. Nevertheless, many countries implemented programmes to give additional support to students, some including psychological and socio-emotional measures (OECD, 2022). This leads us to examine whether the performance of the pandemic cohort (2022 assessment) has suffered across countries.

2. Related work

Many studies employ multiple waves of PISA data to analyse long-term changes and equity patterns. International large-scale tests like PISA are commonly used to evaluate trends in student performance and cross-country discrepancies (OECD, 2022, 2023; OECD, 2023). Low achievement and underperformance have also been the focus of recent research, which uses PISA 2018 data to establish profiles of students with the lowest scores and investigate how school, family, and personal traits influence underachievement across nations (Gutiérrez-de-Rozas, López-Martín, & Carpintero Molina, 2022). Simultaneously, a wide body of research examined how COVID-19 affected learning outcomes, frequently

combining PISA 2018 and 2022 to measure learning loss associated with the pandemic and reveal significant losses in reading and math performance (OECD, 2022; OECD, 2023).

Our project makes a contribution by focusing on low achievement rates rather than average scores, tracing long-term trends (2006–2022), cross-country heterogeneity, and the particular evolution of low achievement between the pre-pandemic (2018) and pandemic cohort (2022; Eurostat, 2023) using a transparent Python-based pipeline on Eurostat/OECD data.

3. Methods

We used the dataset "Low-achieving 15 year olds in reading, mathematics or science" based on data of the OECD with the online data code sdg_04_40 and DOI (https://doi.org/10.2908/SDG_04_40), last updated on 14 December 2023, to answer our research questions presented in the introduction. We worked with Python in VisualStudioCode. Our repository is accessible via: https://github.com/whatamIevendoing-213/PISAProject. In order to work on our research questions we followed the following pipeline:

Data Collection

1. Acquiring the Data
   We downloaded the raw, above mentioned dataset from Eurostat.

Data Cleaning

Since unchecked errors and irrelevant information can skew results and compromise reliability, data cleaning is widely acknowledged as a vital step in the analytics pipeline. According to one source, analysts "spend anywhere from 60–80% of their time carrying out data cleaning tasks," therefore "if your data has inconsistencies or flaws, you can expect that your results will be incorrect, too" (Toʻxtasinov & Sodiqov, 2025). Therefore, before modeling, we eliminated unnecessary metadata, narrowed the dataset to pertinent years, and specifically examined any missing values in the primary outcome variable.

The following steps were followed:

1. *Loading the Data*
   Pandas were imported and the extracted CSV is read into a DataFrame.
2. *Dropping Irrelevant Columns*
   Irrelevant meta data columns like DATAFLOW, LAST UPDATE, freq, sex, unit, CONF_STATUS were removed to keep only meaningful analytical fields in the dataframe.
3. *Filtering by Year and dropping records before 2006*
   Saved records with scores on or after 2006 and removed the rest. We only kept rows where TIME_PERIOD >= 2006. This reduces the dataset to recent years.
4. *Inspecting and extracting Missing Values in OBS_VALUE*

Checked which records have missing values in the main metric (OBS_VALUE). Then, we stored these rows (with columns: field, geo, TIME_PERIOD, OBS_VALUE) into a new list called countries_null.
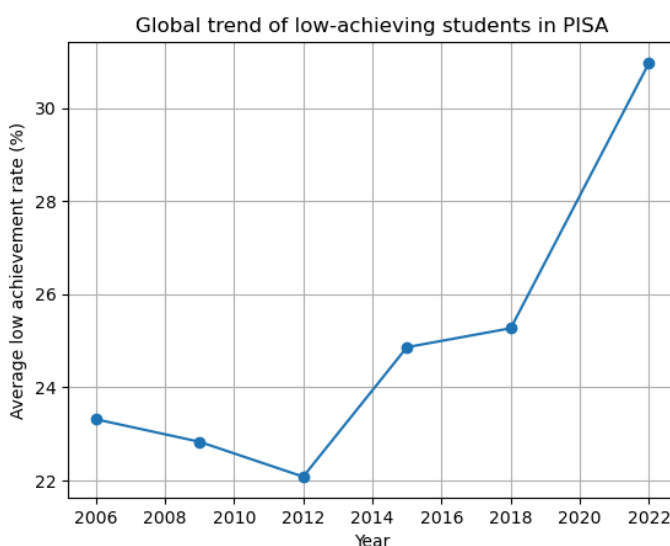
Preprocessing & Visualisation

1. We first analysed the overall time development of low achievement across all countries by computing the yearly mean of low-achievement rates (averaged across reading, mathematics, and science) and visualising it in a global trend plot. This showed a slight decline until 2012, followed by a strong increase up to 2022.

2. We then examined subject-specific trends by calculating yearly averages separately for reading, mathematics, and science. The results revealed that low achievement was consistently highest in mathematics and lowest in science, with all subjects exhibiting a pronounced increase after 2018.

3. Next, we analysed country-level developments by aggregating subject scores into a single country–year average and plotting long-term trends for selected countries (Germany, France, Italy, Spain, and Finland). This revealed substantial heterogeneity in national trajectories.

4. To quantify long-term changes across all countries, we computed first–last differences in low-achievement rates and ranked countries by improvement and deterioration. This showed that most countries experienced increasing low achievement, while Türkiye, Italy, and Romania exhibited the strongest long-term improvements.

5. Next, we visualised the average low-achievement rates in 2006 and 2022 individually, later uniting them in one visualisation to see different trends across countries. We also visualised a boxplot to see the distribution of average rates apart from the average.

6. To work on RQ3 we then computed the changes between the assessment in 2018 and 2022, based on which we visualised a horizontal bar chart.
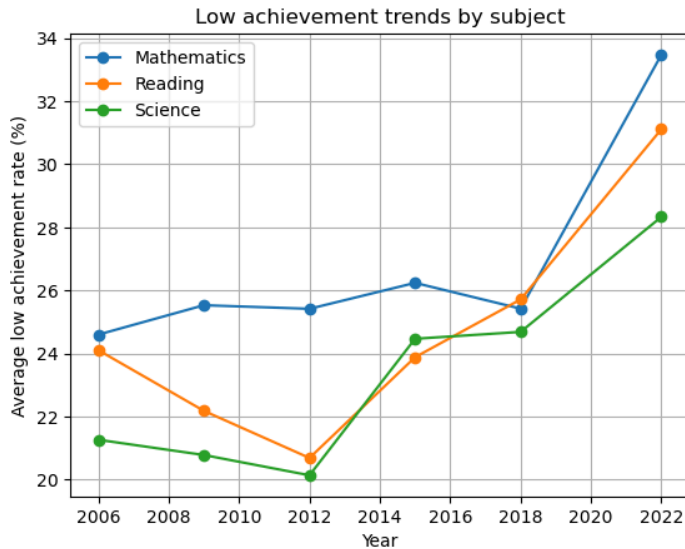
4. Discussion

In the following, we will discuss the research questions based on the visualisations and tables we created with the help of Python.
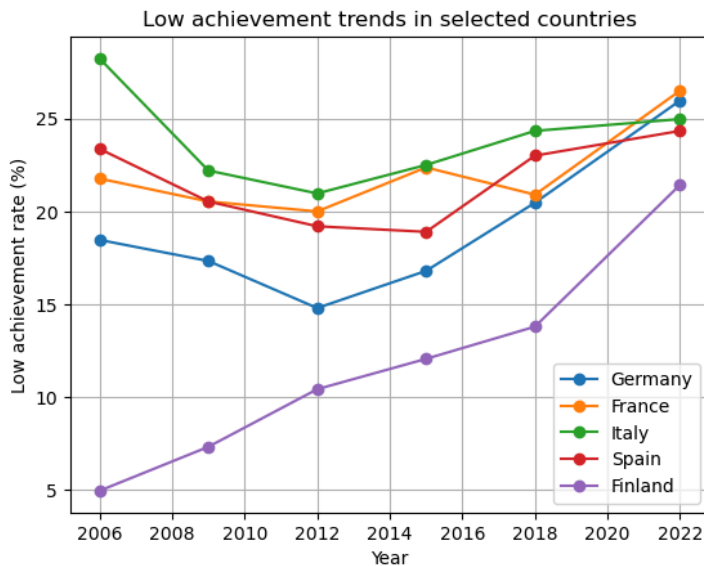
RQ1: How have low achievement rates evolved over time across countries?

This plot shows the global trend of low-achieving students in PISA over the years 2006 - 2022. It shows that there is a decrease in the low achievement rate until increasing sharply in 2012 - 2022.

Low achievement trends by subject

This graph looks at potential differences per subject. It can be seen how low achievement rates were the lowest in Science and the highest in Mathematics. Moreover, we can see that Mathematics' low achievement levels remained comparatively stable until 2018, when they followed the increasing trend with Reading and Science.

Low achievement trends in selected countries

This plot looks at the development of low achieving rates (the average of all three subjects) in selected countries. Here it can be seen how Finland's education system seemed to hold low achievement rates to 5% but still drastically increased until 2022. In contrast, Italy was able to lower its low achievement rate over the years.

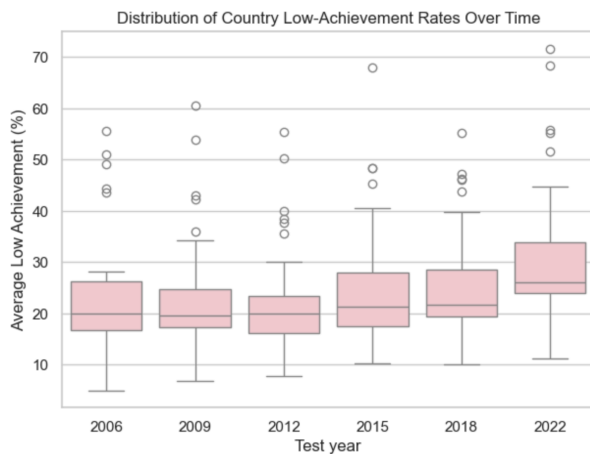|	geo	first	last	change
14	Hungary	21.2	29.5	8.3
37	Sweden	18.4	27.2	8.8
27	Norway	22.4	31.5	9.1
41	United States	24.4	33.9	9.5
5	Cyprus	42.0	53.2	11.2
10	Finland	6.0	18.0	12.0
33	Slovakia	21.0	35.4	14.4
13	Greece	32.5	47.2	14.7
25	Netherlands	11.6	27.4	15.8
15	Iceland	16.9	39.7	22.8

This table shows that most countries experienced an increase in low achievement over time. Iceland, the Netherlands, Greece, and Slovakia exhibit the strongest deteriorations.

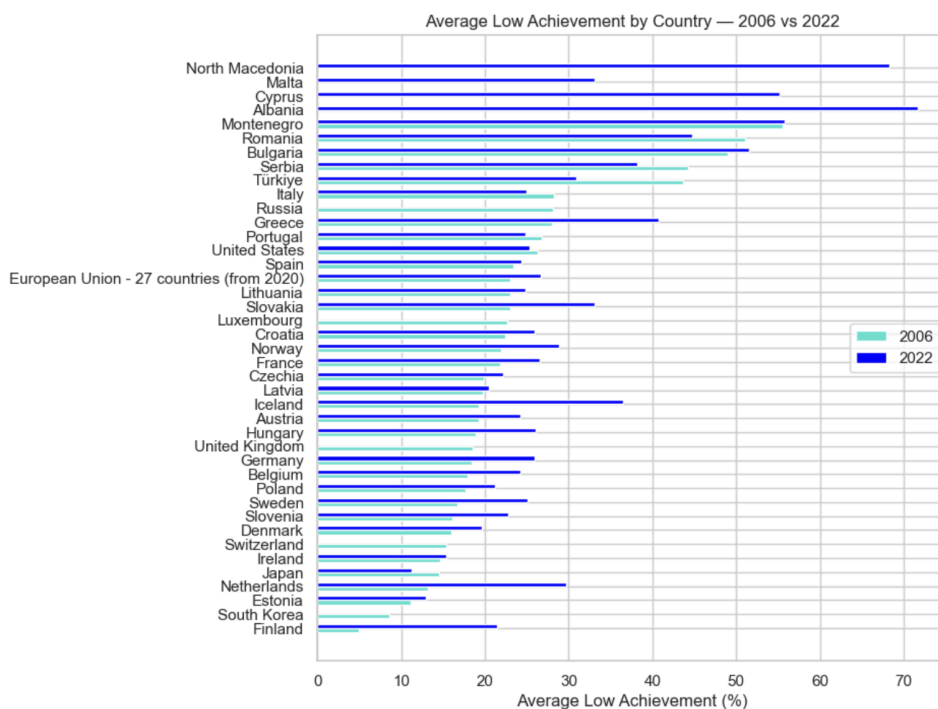|		geo	first	last	change
39		Türkiye	52.2	29.3	-22.9
17		Italy	33.0	21.4	-11.6
30		Romania	52.9	41.7	-11.2
29		Portugal	30.9	21.8	-9.1
32		Serbia	42.6	36.4	-6.2
3		Bulgaria	53.4	48.0	-5.4
18		Japan	13.1	8.0	-5.1
31		Russia	26.8	22.1	-4.7
26		North Macedonia	70.7	66.2	-4.5
19		Latvia	20.8	16.5	-4.3
36		Spain	24.8	21.3	-3.5
28		Poland	20.0	18.6	-1.4
16		Ireland	16.5	15.6	-0.9
24		Montenegro	60.2	59.5	-0.7
0		Albania	67.7	67.4	-0.3

From this table we can see the countries with the largest decrease in low achievement rates. Notably, Turkiye, Italy and Romania experienced a significant decrease.

The analysis shows that low-achievement rates among 15-year-olds have increased substantially over time across most participating countries. While a slight decline was observed between 2006 and 2012, this trend reversed thereafter, with a strong and consistent rise in low achievement up to 2022. This increase is visible both in the global average and across most individual countries. Subject-specific results further indicate that mathematics consistently exhibits the highest share of low-achieving students, while science shows the lowest, although all three domains follow a similar upward trajectory in recent years. Overall, these findings demonstrate a general deterioration in basic educational performance over time and provide strong evidence of substantial cross-country differences in both performance levels and long-term developments.

RQ2: Are there major differences in education performance and its development among countries?

This boxplot confirms the already generally observed negative trend in the development of the average low-achievement rate, especially from 2012 to 2022 on the aggregate level. It however also shows big differences between the assessed countries, as we can see from the numerous and widely spread outliers. The boxplot however does not allow us yet to examine/understand differences in the trajectories of individual countries.

This horizontal bar chart compares the differences in average performance and change of the low achievement rate comparing 2006 to 2022. Some countries display only one bar as they only participated in the 2006 (South Korea, Switzerland, United Kingdom, Luxembourg, Russia) or 2022 (Albania, Cyprus, Malta, North Macedonia) assessment. First of all, we can see in this graph major differences in the education performance across countries in both years, 2006 and 2022. While Finland in 2006 had a low-achievement rate of only 5%, the worst performing country Montenegro saw a 10 times higher rate of low performers. Despite a small improvement, Montenegro in 2022 again was the worst-performing country among those that participated in both the 2006 and 2022 assessments, compared to Japan at 11,4%. This graph also allows us to detect major differences in the development of low-achievement rates across countries. For the majority of countries it reflects the general trend of increases in the low-achievement rates, however while the low achievement rates multiply in some countries (e.g. Finland to 21,4%, Netherlands from 13,2% to 29,7%), they increased much less significantly in many other countries (e.g. Germany from 18,5% to 25,9%, Spain from 23,4% to 24,3%). This contrasts to the countries whose development went into the other direction as their low-achievement rate decreased comparing 2006 to 2022, namely Romania, Serbia, Türkiye, Italy, Portugal, United States and Japan. Overall, this shows that there are

major differences in education performance in single years as well as in the development across time across the assessed countries.

RQ3: Has performance of the pandemic cohort (2022 assessment) suffered across countries?

Number of countries that worsened	32
Number of countries that improved	4
Mean change across all countries	5.057407407407407

For this table we computed the change in average low achievement rates across countries between 2018 and 2022. The table provides a good first overview of the general trend of increasing low achievement rates ("Number of countries that worsened"). While we computed a mean, we need to run further calculations to get a better idea of individual changes.

Countries with largest increase in low achievement (worst hit)

	geo	change
0	Albania	24.466667
5	Cyprus	15.333333
26	North Macedonia	13.166667
15	Iceland	12.533333
25	Netherlands	9.800000

Countries with decrease in low achievement (improved)

	geo	change
18	Japan	-1.766667
32	Serbia	-0.366667
23	Malta	-0.133333
4	Croatia	-0.066667
16	Ireland	0.500000

These tables show the five countries with the largest increases and decreases in low achievement rates comparing 2018 to 2022. The biggest increase in low achievement rates (therefore the greatest decline in test results) could be identified in Albania, where the average low-achievement rate increased by 24,5 percentage points, followed by Cyprus

(15,3) and North Macedonia (13,2). Some countries managed to improve their performance despite the pandemic, although only slightly: The low achievement rate in Japan decreased by -1,7 percentage points, followed by Serbia (-0,4) and Malta (-0,13). These positive changes however are by far not as significant as the deteriorations in achievement.

This horizontal bar chart shows the development of the average low achievement rate from 2018 to 2022 in 34 countries, the 2018 values being the "baseline" at 0 and the bars indicating the increase/decrease in the low achievement rate. It can again be seen that the low achievement rate increased in the majority of countries, to varying degrees.

The analysis shows that low-achievement rates among 15-year-olds of the pandemic cohort (comparing 2018 to 2022) suffered across the majority of countries, however not all. Out of 36 countries that participated in both assessments in 2018 and 2022, the performance of 32 worsened and the performance of 4 improved. The median change was around 5 percentage points. Overall, these findings demonstrate a global deterioration in educational performance in the pandemic cohort (2022 assessment), with a few countries being the exception.

5. Conclusion

This paper examined trends in low-achievement among 15-year olds across different countries from 2006 to 2022, focusing on long-term trends and cross-country-differences as well as the impact of the COVID-19 pandemic. The results show that low-achievement rates increased notably since 2012. Substantial differences were found across countries, both in performance levels and trajectories over time. While most countries experienced increasing low-achievement rates, a smaller group including Türkiye, Italy and Romania saw notable improvements, whereas several traditionally high-performing systems deteriorated. The analysis provides evidence that the pandemic cohort was negatively affected in most countries, with only few exceptions. Overall, the findings highlight the need for education policies that strengthen basic skills and increase the resilience of the education system towards shocks.

6. References

Coughlan, Sean. 2013. How PISA Became the World's most Important Exam. *BBC News*. Available online: https://www.bbc.com/news/business-24988343 (accessed on 16 December 2025).

Eurostat (2023), [sdg_04_40] Low achieving 15-year-olds in reading, mathematics or science. https://ec.europa.eu/eurostat/databrowser/view/sdg_04_40/default/table?lang=en&category=sdg.sdg_04

Gutiérrez-de-Rozas, B., López-Martín, E., & Carpintero Molina, E. (2022). Defining the profile of students with low academic achievement: A cross-country analysis through PISA 2018 data. Frontiers in Education, 7, 910039. https://doi.org/10.3389/feduc.2022.910039

OECD. (2022, October 3). Education at a Glance 2022. OECD. https://www.oecd.org/en/publications/education-at-a-glance-2022_3197152b-en.html

OECD. (2023, September 12). Education at a Glance 2023. OECD. https://www.oecd.org/en/publications/2023/09/education-at-a-glance-2023_581c9602.html

OECD (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education, PISA. OECD Publishing, Paris. https://doi.org/10.1787/53f23881-en.

To'xtasinov, A. I., & Sodiqov, V. S. (2025). The importance of data cleaning in data analysis. *Journal of New Century Innovations, 78*(2), 115–117. https://scientific-jl.com/new/article/view/19541